DOI:
10.1016/j.neubiorev.2018.03.002

Document Version
Peer reviewed version

Title: Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants

Authors: Miquel A. Fullana, Anton Albajes-Eizagirre, Carles Soriano-Mas, Bram Vervliet, Narcís Cardoner, Olívia Benet, Joaquim Radua, Ben J. Harrison

PII: S0149-7634(17)30960-0
DOI: https://doi.org/10.1016/j.neubiorev.2018.03.002
Reference: NBR 3060

To appear in:

Received date: 21-12-2017
Revised date: 27-2-2018
Accepted date: 3-3-2018

Please cite this article as: Fullana MA, Albajes-Eizagirre A, Soriano-Mas C, Vervliet B, Cardoner N, Benet O, Radua J, Harrison BJ, Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants, *Neuroscience and Biobehavioral Reviews* (2010), https://doi.org/10.1016/j.neubiorev.2018.03.002

Fear extinction in the human brain: a meta-analysis of fMRI studies in healthy participants

Miquel A. Fullana[1,2,*], PhD; Anton Albajes-Eizagirre[1],MSc; Carles Soriano-Mas[3,4], PhD ; BramVervliet[5,6], PhD; Narcís Cardoner[2,7], MD, PhD; Olívia Benet[8], BSc; Joaquim Radua[1,9,10], MD, PhD; and Ben J. Harrison [11,*], PhD.

[1] FIDMAG Germanes Hospitalàries, CIBERSAM, Sant Boi de Llobregat, Barcelona, Spain; [2] Department of Psychiatry, Universitat Autònoma de Barcelona, Barcelona, Spain; [3] Department of Psychiatry, Bellvitge University Hospital-IDIBELL, CIBERSAM, Barcelona, Spain; [4] Department of Psychobiology and Methodology of Health Sciences, Universitat Autònoma de Barcelona, Spain; [5]Center for Excellence on Generalization in Health and Psychopathology, University of KU Leuven, Leuven, Belgium; [6] Department of Psychiatry, Harvard Medical School, Boston, US & Department of Psychiatry, Massachusetts General Hospital, Boston, US; [7]Depression and Anxiety Unit, Mental Health Department, CIBERSAM, Parc Taulí Sabadell University Hospital, Barcelona, Spain; [8]Autonomous University of Barcelona, Spain; [9] Institute of Psychiatry, King's College London, De Crespigny Park, London,UK; [10] Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; [11] Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Victoria, Australia.

*These authors contributed equally to this work

Authors for correspondence: Dr. Miquel A. Fullana. FIDMAG Germanes Hospitalàries, CIBERSAM, Sant Boi de Llobregat, Barcelona, Spain. (0034) 93 5811223, miguelangelfullana@gmail.com. A/Prof. Ben J. Harrison, Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Level 3, 161 Barry Street, Melbourne, Victoria, Australia, habj@unimelb.edu.au.

Highlights

- Reports a set of meta-analyses of human fMRI fear extinction studies involving over 1300 participants
- Human fear *extinction learning* and *extinction recall* are consistently distinct in their neural correlates
- The contribution of prefrontal cortical subregions to extinction appears to be more nuanced than what is suggested by current translational models

ABSTRACT

The study of fear extinction represents an important example of translational neuroscience in psychiatry and promises to improve the understanding and treatment of anxiety and fear-related disorders. We present the results of a set of meta-analyses of human fear extinction studies in healthy participants, conducted with functional magnetic resonance imaging (fMRI) and reporting whole-brain results. Meta-analyses of fear extinction learning primarily implicate consistent activation of brain regions linked to threat appraisal and experience, including the dorsal anterior cingulate and anterior insular cortices. An overlapping anatomical result was obtained from the meta-analysis of extinction recall studies, except when studies directly compared an extinguished threat stimulus to an unextinguished threat stimulus (instead of a safety stimulus). In this latter instance, more consistent activation was observed in dorsolateral and ventromedial prefrontal cortex regions, together with other areas including the hippocampus. While our results partially support the notion of a shared neuroanatomy between human and rodent models of extinction processes, they also encourage an expanded account of the neural basis of human fear extinction.

Introduction

Just as learning to predict threat (*fear* or *threat conditioning*) is critical to one's survival and wellbeing, learning that a prior threat no longer signals danger (*fear* or *threat extinction*) is also fundamentally adaptive. Recognized almost a century ago (Pavlov, 1929), the construct of 'fear extinction' has received renewed interest over the past two decades, particularly with regard to its underlying neural basis (Dunsmoor et al., 2015; Myers and Davis, 2002; Quirk and Mueller, 2008; Tovote et al., 2015). It is expected that a deeper knowledge of fear extinction mechanisms

will increase our understanding of anxiety and fear-related disorders, to which extinction deficits have been broadly implicated (Duits et al., 2015). Similarly, because fear extinction principles form the basis of successful exposure therapies for these disorders, it is also hoped that their ongoing neuroscientific study will lead to further treatment advances. For example, a recent study confirmed that brain activation during fear extinction learning predicted exposure therapy outcome in socially anxious individuals (Ball et al., 2017).

Importantly, detailed neurocircuitry accounts of extinction processes based on rodent studies have been broadly anatomically translated to human studies via the application of functional magnetic resonance imaging (fMRI)(Dejean et al., 2015; Hartley and Phelps, 2010; Linnman et al., 2012; Milad and Quirk, 2012). This work has led to some consensus about the brain's core 'fear extinction network', which typically includes the amygdala, hippocampus and ventromedial prefrontal cortex (vmPFC)(Gottfried and Dolan, 2004; Hauner et al., 2013; Kalisch, 2006; Milad et al., 2007; Phelps et al., 2004). However, as more fMRI studies have targeted this network, a somewhat less consistent picture has emerged. For example, during fear extinction learning, increased and decreased amygdala responses to conditioned (CS+) versus non-conditioned stimuli (CS-) have been reported (Gottfried and Dolan, 2004; LaBar et al., 1998; Milad et al., 2007; Sehlmeyer et al., 2011). The vmPFC also demonstrates a complex activity profile that challenges straightforward interpretation (Harrison et al., 2017). Relative to a resting baseline, vmPFC activity is initially suppressed by CS+ versus CS- during conditioning; a difference that gradually minimizes over the course of extinction learning as CS+ become less threatening, or more safe (Milad et al., 2007; Phelps et al., 2004; Schiller et al., 2008; Schiller and Delgado, 2010). Somewhat more consistent vmPFC responses are observed during fear extinction recall (Kalisch, 2006; Milad et al., 2007; Phelps et al., 2004), which parallels the

findings of some (Quirk and Mueller, 2008) but not other animal studies (Bukalo et al., 2015; , Do-Monte et al., 2015). Indeed, with regards to the latter, there is current debate regading the precise contribution of the vmPFC to extinction learning vs. recall processes, which accordingly, present a challenge to human fear extinction models that emphasise a more primary role in extinction recall (Clem and Schiller, 2016; Delgado et al., 2016).

In a previous meta-analysis using the activation likelihood estimation (ALE) method (Eickhoff et al., 2012) and including 10 studies and 154 participants, anterior and posterior vmPFC subregions, as well as the dorsomedial PFC, were identified as consistently activated in fMRI fear extinction studies (Diekhof et al., 2011). However, this meta-analysis combined extinction learning (where most often a CS+ versus a CS- is compared) and extinction recall (where often an extinguished versus an unextinguished CS+ are compared) studies, which complicates its interpretation. Moreover, this meta-analysis did not explore the effects of experimental variables on extinction or assess the robustness of findings across studies. Regarding extinction recall, Menz et al. (2016) recently conducted an ALE-based meta-analysis that focused on vmPFC activity across 15 studies. While confirming a role for the vmPFC, the contribution of other brain areas to extinction recall was not examined. This analysis also grouped together studies employing different contextual manipulations to assess extinction recall, which may have obscured its findings, given that extinction is known to be highly context-dependent (Bouton, 1993; Maren et al., 2013). It also pooled data from studies of healthy 'control' and trauma-exposed 'control' participants, which may not be compatible populations for studying neural fear extinction processes (Marin et al., 2016).

Our current aim was therefore to address such limitations and to provide an updated and extended meta-analysis of fMRI fear extinction learning and recall studies, including a total of

more than 1300 participants. We investigated the role of several experimental variables in extinction learning and conducted, for the first time, a direct meta-analytic comparison of brain activation patterns evoked during fear conditioning versus fear extinction learning. Our analyses were primarily based on the inclusion of original whole-brain statistical maps, which increased our sensitivity to identify the most robust brain activation effects across studies.

Methods and materials

We followed MOOSE guidelines for meta-analyses of observational studies (Stroup et al., 2000)**.**

Search and inclusion of studies

A comprehensive literature search using PubMed, Web of Knowledge and Scopus was conducted for English-language peer-reviewed studies of conditioned fear extinction (extinction learning and extinction recall) in human healthy adults [1] (age > 18 years) through August 31, 2017 (See PRISMA diagrams in **Supplementary Figures S1 and S2**). The search terms were: 'fMRI' or 'magnetic resonance imaging', 'fear', 'extinction', and their combinations. Returned articles were also manually inspected for additional studies. We focused on studies that assessed fear extinction using delay differential cue-conditioning paradigms (i.e, where a CS+ and a CS- are presented and the CS+ precedes the US) and that reported direct comparisons between a CS+ and CS- during fear extinction learning or fear extinction recall (where some studies compared an extinguished versus an unextinguished CS+; see below). If pharmacological or other challenges were involved, only results from the placebo/non-challenge condition were included. For extinction learning meta-analyses, studies were excluded if they did not provide evidence of successful initial fear conditioning (e.g., increased skin conductance response (SCR) to CS+ >

[1] We excluded "trauma-exposed" controls from post-traumatic stress disorder studies (see Introduction)

CS-); if they used masked CSs or an unconditioned stimulus (US) with ambiguous meaning; or if the CS–US contingencies changed during conditioning. For extinction recall meta-analyses, studies were excluded if they did not provide evidence of successful extinction learning (e.g., reduced differential SCR to CS+ > CS-). There was no duplication of participant groups across studies. We contacted authors regarding their willingness to share original maps, or to provide whole brain analysis results if these maps were not available.

Extinction learning meta-analyses

We first conducted a meta-analysis of all extinction learning studies (**Extinction learning meta-analysis**). In certain studies, all CSs trials during extinction were included in the analysis, whereas in others, 'early' and 'late' extinction phases were modeled separately. When more than one contrast was available from a given study, we sought to include the contrast involving all trials. If this contrast was not available, we focused on late extinction learning trials (Milad et al., 2007). The influence of gender (% female), age (sample mean of participants per study), reinforcement rate and number of CSs during conditioning, and number of CSs during extinction was examined via meta-regression. Moreover, we directly compared the extinction phase of studies where conditioning and extinction took place in a different context versus studies where conditioning and extinction took place in the same context (**Different context versus same context**). To decrease heterogeneity, we excluded from this analysis studies assessing context and cue conditioning in the same experiment, studies using instructed conditioning or extinction, and studies investigating delayed extinction. Finally, we also directly compared the studies (n=27, 677 participants) included in our previous meta-analysis of fear conditioning (Fullana et al., 2016) with the fear extinction learning studies included in the current meta-analysis (**Fear**

conditioning versus extinction learning), with the exception of those studies that were included in the former conditioning meta-analysis. We focused on results corresponding to the overall assessment of conditioning and extinction trials (CS+ > CS-), respectively. If these contrasts were not available, we included results corresponding to the analysis of 'early' phase conditioning and 'late' phase extinction trials (CS+ > CS-), as there is some consensus that the neural correlates of each process are best captured during these respective phases (Fullana et al., 2016; Milad et al., 2007).

Extinction recall meta-analyses

We conducted a meta-analysis of all extinction recall studies (**Extinction recall meta-analysis**). Additionally, we conducted two separate meta-analyses including 1) only studies where extinction recall was tested in the *same context* as extinction learning and included at least two CS+ during extinction recall (an extinguished CS+, or 'CS+E', and an un-extinguished CS+, or 'CS+U'; thus corresponding to the primary contrast of CS+E > CS+U) (**Extinction recall contrast 1**); 2) studies where extinction recall was tested using a CS+ > CS- contrast (**Extinction recall contrast 2**).

Meta-analytic approach

The anisotropic effect-size version of seed-based *d* mapping software (ES-SDM; http://www.sdmproject.com; Radua et al., 2012) was used to generate voxel-wise (random effects) activation effect size maps corresponding to the aforementioned analyses and contrasts of interest. ES-SDM is a neuroimaging meta-analytic approach that is capable of combining tabulated brain activation results (i.e., regional peak statistic and coordinate information) with

actual empirical voxel-wise 'activation maps' (e.g., statistical parametric maps). We focused on the comparison CS+ > CS- in all extinction learning meta-analyses. For extinction recall, we used both the CS+ > CS- and the CS+E > CS+U contrasts.

To assess the robustness of the main findings, we conducted jackknife sensitivity analyses (to check for replicability) and we also used the I^2 index and Egger's test (Egger et al., 1997) to assess the heterogeneity of effect sizes and publication biases, respectively. Statistical significance was assessed using a randomization test and ES-SDM default thresholds (voxel-level P< 0.005 uncorrected, minimum cluster extent 10 contiguous voxels). Previous simulations indicate that this threshold approximates a control for multiple comparisons and provides an optimal balance between sensitivity and false-positive rate (Radua et al., 2014). A more conservative threshold (P< 0.0005) was applied to the meta-regression analysis to familywise control for the multiple testing (several meta-regressions).

Results

Meta-analysis characteristics

The number of studies, participants, and main characteristics of the studies included in each meta-analysis are reported in **Tables 1, 2,** and **S1.** Peaks and coordinates for studies for which activation maps were not available are reported in tables **S2** to **S5.**

---Table 1---

---Table 2---

Extinction learning meta-analysis (31 studies; 1074 participants).

Figure 1 presents the meta-analytic mean map of brain regions consistently activated during extinction learning. These regions included the rostro-dorsal anterior cingulate cortex extending to the pre-supplementary motor cortex and medial prefrontal cortex, bilateral anterior insular cortex extending to frontal operculum, dorsolateral prefrontal cortex, anterior putamen extending to ventral caudate, bilateral ventral pallidum, anterior and medial thalamus, and midbrain/dorsal pons (~ periaqueductal grey). Robustness analyses indicated that these findings were preserved in most studies, and there was no evidence of heterogeneity or publication bias (except for the clusters including the left ventral caudate and the right premotor cortex) (**Table S6**). As a supplement to these results, we also performed a restricted analysis of studies reporting late extinction phase effects alone[2]. Results are presented in **Table S7** and demonstrate that the activated regions most consistently observed in late extinction phases included the anterior thalamus, ventral putamen and right anterior insular cortex.

No significant associations via meta-regression were identified between gender, age, reinforcement rate or number of CSs during conditioning, or number of CSs during extinction and the fear extinction activation effect.

---Figure 1---

Different context versus same context (8 studies, 158 participants vs 17 studies, 635 participants). The direct comparison of studies where extinction learning took place in different (versus the same) context as fear conditioning identified significantly greater activation of the bilateral middle occipital cortex and left somatosensory-supramarginal cortex. The direct comparison of

[2] It was not possible to conduct a formal meta-analysis comparing "early" and "late" extinction phases due to the limited number of individual studies calculating such contrast.

studies where extinction learning took place in the same (versus a different) context to that as fear conditioning identified significantly greater activation of the left anterior insular cortex, right posterior insular cortex and anterior cerebellum (see **Table S8**).

Fear conditioning versus extinction learning (27 studies, 677 participants vs 24 studies, 834 participants).

The direct comparison of fear conditioning versus extinction learning identified consistently greater activation of the midline supplementary motor area extending to dorsal anterior cingulate cortex, anterior and mid insular cortex, ventral caudate nucleus, lower and upper brainstem regions, including the periaqueductal gray, and primary and secondary somatosensory cortex during fear conditioning (**Figure 2; Table S9**). No significant differences in activation were observed from the direct comparison of extinction learning to fear conditioning.

---Figure 2---

Extinction recall meta-analyses (16 studies, 342 participants).

Figure 3A presents the meta-analytic mean map of brain regions consistently activated during all extinction recall studies (see **Table S10**). These regions included the dorsal anterior cingulate cortex, bilateral anterior insular cortex extending to frontal operculum, the right septal-hypothalamic region, and left second somatosensory-parietal opercular cortex. Further analysis of extinction recall contrast types indicated that this overall effect was mainly driven by studies employing the CS+ vs. CS- contrast (**Figure 3B**, **Table S11**). The meta-analysis of studies comparing CS+E to CS+U instead identified significant activation of the left anterior and right

dorsolateral prefrontal cortex, the subgenual cingulate-posterior ventromedial prefrontal cortex, left lateral orbitofrontal cortex, left parietal operculum, and right anterior hippocampus (**Figure 4**, **Table S12**). All findings from the extinction recall meta-analyses were replicable and there was no evidence of heterogeneity or publication bias.

---Figure 3---

---Figure 4---

Discussion

We have conducted an updated and extended meta-analysis to examine the consistency of fMRI studies in evoking activation of core 'fear extinction network' regions. While our results partially support the notion of a shared neuroanatomy between human and rodent models of extinction processes, they also encourage an expanded account of the neural basis of human fear extinction.

Unlike early fMRI studies (Gottfried and Dolan, 2004; LaBar et al., 1998; Milad et al., 2007; Phelps et al., 2004), the current meta-analyses did not identify consistent amygdala involvement during fear extinction learning. This observation is not entirely surprising: the precise nature of amygdala activity varied across early studies, including reports of both increased and decreased responses to CS+ versus CS-, as well as preferential responding to both early and late extinction trials. The absence of consistent amygdala involvement during extinction learning (including *late* extinction) parallels the findings of our recent meta-analysis of fear conditioning, which also separately examined early versus late trials (Fullana et al., 2016). Interpretively, we favor the notion that while human amygdala activity is involved in associative fear/threat learning processes via its central role in species-conserved 'defense-

survival' circuits (LeDoux, 2014), the engagement of these circuits may not be reliably indexed in conventional fMRI fear learning experiments, as compared to animal studies. This result may be due to the limited anatomical resolution of fMRI. For example, rodent studies have shown that some subnuclei of the amygdala may show increased activity and others show decreased activity during fear extinction (Repa et al., 2001; see Herry et al., 2010). Other factors that may explain such results are the standard design of fMRI experiments, including the necessary averaging across multiple trial repetitions; the use of different 'baselines' (see further); and the *sparse* (Bach et al., 2011) nature of amygdala activity evoked by these tasks in humans. Regarding the latter, animal studies suggest that it is implausible that consistent CS+>CS- responses are detected in the amygdala during fear learning studies using mass-univariate approaches (Ciocchi et al., 2010; Reijmers et al., 2007) and that other methods such as multivariate fMRI (Bach et al., 2011) may be better suited to this end. Finally, it is also possible that the amygdala, as well as other brain areas (see further) are more prominently recruited during intense fear states, such as those experienced in clinically anxious populations, and therefore were less evident in our meta-analysis focusing on fear extinction in healthy individuals.

Another finding from early studies that was not reproduced relates to the involvement of the vmPFC in extinction learning, including late extinction learning (Gottfried and Dolan, 2004; Milad et al., 2007; Phelps et al., 2004). There is mounting evidence that the vmPFC preferentially responds to safety (CS-) versus threat (CS+) signals in human fear conditioning studies (Harrison et al., 2017; Schiller and Delgado, 2010). During extinction learning, the previous CS+ takes on the general properties of a safety signal (CS-), resulting in the standard contrast of CS+ > CS- responses being biased toward characterizing minimal difference between

them (safety vs. safety). This design issue is further complicated by the fact that studies have often relied upon additional comparisons to a 'neutral' baseline state in order to illustrate differential vmPFC activity changes. Such baselines have been derived from 'rest/non-task' intervals interleaved between CS trial presentations, during which participants typically view a central fixation stimulus (e.g. white cross). Two factors complicate their use: firstly, vmPFC activity is characteristically *high* during such non-task versus task states: a phenomenon linked to the concept of the 'default mode network' (Harrison et al., 2011, 2008; Raichle et al., 2001). Secondly, it cannot be ruled out that these intervals/stimuli also take on safety signal properties themselves, which would additionally modulate vmPFC activity (Harrison et al., 2017). If these issues are successfully addressed, we expect that future studies may characterize more robust involvement of the vmPFC in extinction learning, as noted in some of the more recent animal studies (e.g. Do Monte et al., 2015).

Extinction learning was instead linked to a pattern of brain activation that is more reminiscent of fear conditioning (Fullana et al., 2016; Mechias et al., 2010). In conditioning studies, participants consistently demonstrate activation of brain 'central autonomic network' regions, including its main cortical components – the anterior insular and dorsal anterior cingulate cortices (Fullana et al., 2016). Co-activation of these regions is frequently observed in human fMRI studies and has been linked prominently to the elicitation of negative affective states, including threat-related anticipatory anxiety (Etkin et al., 2011; Medford and Critchley, 2010). One hypothesis is that these regions contribute fundamentally to the subjective experience of anxious and fearful states, particularly at the level of interoceptive (bodily) awareness (Fullana et al., 2016; Harrison et al., 2015). Considering that fear responses to CS+ vs. CS- are rarely

completely diminished during extinction learning, it seems reasonable that this contrast reliably maps enduring activation of these brain regions.

When directly compared to fear conditioning, brain activations observed during extinction learning were also significantly less robust (**Figure 2**). We know that extinction learning is a *fragile* phenomenon (Bouton, 2002; Giustino and Maren, 2015; Pavlov, 1929). Such fragility could arise from the competition between two memories (a "conditioning" versus an "extinction" memory) and the neural data presented here could be the result of such competition. However, there are potential alternative interpretations. It is possible that the same network of brain regions evoked during fear conditioning perform a different function during fear extinction learning (see Hermans et al., 2006) or that the reduced CS+ vs CS- activity during extinction learning represents the degradation of CS-US associations after repeated presentations of the CS+ without reinforcement (see Rescorla and Wagner, 1972). A further possibility is that the enduring neural threat response observed during extinction learning reflects an important dissociation between fear measures (Hugdahl, 1980); that is, neural responses may reflect an intact CS-US association that is, however, no longer expressed at a behavioral level. It is also possible that CS-elicited neural responses during extinction learning provide little information about *the precise mechanisms* of extinction. For example, according to some authors (Craske et al., 2014), extinction learning relies on expectancy violations (i.e., prediction errors) and it is the *consolidation* of extinction learning what determines how effective extinction ultimately is. A related hypothesis is that neural responses to CS offsets rather than CS onsets may more readily capture such expectancy violations and therefore neural extinction learning mechanisms (Raczka et al., 2011).

Although we were unable to identify brain activity changes that were specific to extinction learning versus conditioning, the meta-analysis of extinction learning studies has highlighted a more prominent involvement of dorsolateral prefrontal cortex than what has been suggested from individual studies. While extinction learning is considered a form of implicit emotion regulation (Schiller and Delgado, 2010), there has been little suggestion in fMRI studies that it also engages prefrontal cortical regions linked with more explicit 'cognitive' forms of emotion regulation (Delgado et al., 2008; Ochsner and Gross, 2005). Nevertheless, behavioral studies suggest that cognitive-regulatory factors may be more prominently involved in human fear extinction learning than conditioning (Lovibond, 2004). These factors may include both 'low level' (e.g., selective attention, episodic memory) and 'higher level' processes (e.g. beliefs and expectancies) (Hermans et al., 2006; Lovibond, 2004; Lovibond and Shanks, 2002). All of these domains have been demonstrated to recruit dorsolateral prefrontal cortical involvement in human neuroimaging studies, and may have contributed to the observed findings here. Although this result will await confirmation, the broader involvement of prefrontal cortical areas does align with some recent models that place greater emphasis on the higher cognitive neural circuits contributions to human fear and anxiety processes (LeDoux and Pine, 2016).

Cerebellar activation also emerged as a consistent observation across the different extinction contrasts, and most prominently in association with extinction learning. As confirmed here, activation of the anterior cerebellum (vermis) has been most strongly emphasized in association to extinction learning in fMRI studies, with some suggestion that it may participate in the autonomic regulatory aspects of fear inhibition (Kattoor et al., 2014; Lange et al., 2015; Utz et al., 2015). While it is difficult to draw further conclusions from the meta-analytic results, they

do confirm that anterior cerebellar activation is a consistent anatomical correlate of the extended neural circuitry of fear extinction learning.

Our meta-analysis on extinction learning focused on the *immediate* (as opposed to *delayed*) effects of extinction after conditioning. Animal research suggests that immediate extinction can modify the consolidation of fear memories (Myers et al., 2006) and human research has shown that immediate extinction can produce a higher fear reduction than delayed extinction (Norrholm et al., 2008), and that early exposure therapy after trauma (equivalent to immediate extinction) reduces fear symptoms (Rothbaum et al., 2012). However, some studies also suggest that immediate extinction results in *less* fear reduction than delayed extinction (see Maren, 2014) and it is possible that both 'types' of extinction involve at least partially different neural mechanisms (LeDoux and Pine, 2016). These issues considered, immediate fear extinction learning (measured at the neural level) has recently been shown to predict exposure therapy outcome (Ball et al., 2017), which may speak to the potential clinical relevance of our findings.

Our extinction recall results highlight the influence that different experimental designs have on characterizing the neural correlates of some fear learning processes. Whereas the CS+ vs CS- comparison elicited a similar pattern to extinction learning, the results of the CS+E vs CS+U contrast were more consistent with past studies emphasizing vmPFC-hippocampal/subcortical circuitry (Kalisch, 2006; Milad et al., 2007), and in particular, a role for the vmPFC in mediating successful fear response inhibition (Dejean et al., 2015; Hartley and Phelps, 2010; Milad and Quirk, 2012). While both experimental approaches are valid (Lonsdorf et al., 2017), the current results suggest that the CS+E vs CS+U contrast better captures the hypothesized vmPFC-subcortical contributions to safety signal processing and fear inhibition/regulation. The results of the latter contrast were, however, more anatomically discrete. As noted in the introduction, the

role of the vmPFC in fear extinction processes is still debated. In fact, recent animal research suggests that the primary contribution of the vmPFC to fear extinction may relate to contextual processing rather than fear inhibition *per se* (Pennington et al., 2017).

Of note, and similar to our extinction learning results, we also found increased activity in the dlPFC during extinction recall, but again only for the CS+U vs CS+E contrast (see Table S12). One possibility is that the dlPFC during extinction recall has a similar role to the vmPFC during extinction learning (i.e., dampening down subcortical responses), given the known role of the dlPFC in emotional regulation (Hartley and Phelps, 2010). Another possibility is that the dlPFC be the main contributor to the retrieval of the extinction memory (versus the fear memory) during recall. These possibilities deserve further research.

The study of extinction recall in humans is challenging and it is likely that small procedural variations yield different outcomes (Lonsdorf et al., 2017). It is possible that the contrast of reinforced vs non-reinforced stimuli and extinguished vs unextinguished CSs represent assessments of different learning processes rather than different assessments of the same process. It is also possible that individual differences play a greater role in extinction recall in comparison to extinction learning (see Gershman and Hartley, 2015; Shumake et al., 2014) and that this variability is also expressed at the neural level.

This study has several limitations. Firstly, it seems likely that the absence of significant meta-regression results for extinction learning may have been influenced by a low degree of variability across studies in the demographic and task parameters that were assessed. Secondly, while we only included studies with reported evidence of conditioning/extinction learning at a behavioral level, the actual strength of learning in each study was not assessed. Finally, there were a limited number of available extinction recall (compared to extinction learning) studies for

the meta-analysis, although we were able to identify significant effects across all meta-analyses when applying an equivalent statistical threshold and corresponding robustness tests.

In conclusion, we have provided a meta-analytic summary of the results of fMRI fear extinction studies. The findings endorse some current accounts of the neurobiological basis of fear extinction that have been translated between animal and human studies, but also suggest that more nuanced accounts of the prefrontal cortical contribution to human fear extinction, in particular, may be possible. To this end, the development of novel and more sophisticated experimental studies, including neuroimaging studies, should continue to drive the field forward in terms of advancing our understanding of extinction-related processes and their contribution to common anxiety and fear-related disorders, including potentially optimized treatments.

CONFLICT OF INTEREST

The authors report no conflict of interest.

ACKNOWLEDGEMENTS

We thank the following authors for providing additional data or information in support of the meta-analyses: Fredrik Åhs; Marta Andreatta; B.J. Casey; Monika Eckstein; Sigrid Elsenbruch; Andrea Hermann; Daphne Holt; Adriane Icenhour; Raffael Kalisch; Joswin Kattoor; Merel Kindt; Tilo Kircher; Manuel Kuhn; Floris Klumpers; Annegret Krause-Utz; Kevin LaBar; Katja Lindner; Clas Linnman; Tina B. Lonsdorf; Manesh Menon; Mareike Menz; Christian Merz; Mohammed Milad; Tanaz Molapour; Jayne E. Morriss; Frauke Nees; Paul Pauli; Robert Scharfenort; Victor Spoormaker; Rudolf Stark; Benjamin Straube; Sonja M. van Well; and Renee M.Visser.

This work was supported by "Miguel Servet" contracts from the Carlos III Health Institute (Spain) to Dr. Soriano-Mas (CP10/00604) and Dr. Radua (CP14/00041); from grants by the same institution and Fondo Europeo de Desarrollo Regional to Dr. Cardoner (PI12/01306; PI12/00273) and to Dr. Fullana (PI12/00273; PI16/00144); from a KU Leuven program funding grant (PF/10/005) and FP7 Marie Curie International Outgoing Fellowship (627743) to Dr. Vervliet. A/Prof. Harrison is supported by a National Health and Medical Research Council of Australia (NHMRC) Clinical Career Development Award (1124472).

Supplementary data.

Supplementary data associated with this article can be found, in the online version.

References

Åhs, F., Kragel, P.A., Zielinski, D.J., Brady, R., Labar, K.S., 2015. Medial prefrontal pathways for the contextual regulation of extinguished fear in humans. Neuroimage 122, 262–271. https://doi.org/10.1016/j.neuroimage.2015.07.051

Bach, D.R., Weiskopf, N., Dolan, R.J., 2011. A stable sparse fear memory trace in human amygdala. J. Neurosci. 31, 9383–9. https://doi.org/10.1523/JNEUROSCI.1524-11.2011

Ball, T.M., Knapp, S.E., Paulus, M.P., Stein, M.B., 2017. Brain activation during fear extinction predicts exposure success. Depress. Anxiety 34, 257–266. https://doi.org/10.1002/da.22583

Benson, S., Kattoor, J., Kullmann, J.S., Hofmann, S., Engler, H., Forsting, M., Gizewski, E.R., Elsenbruch, S., 2014. Towards understanding sex differences in visceral pain: Enhanced reactivation of classically-conditioned fear in healthy women. Neurobiol. Learn. Mem. 109, 113–121. https://doi.org/10.1016/j.nlm.2013.12.014

Bouton, M.E., 2002. Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. Biol. Psychiatry 52, 976–986. https://doi.org/10.1016/S0006-3223(02)01546-9

Bouton, M.E., 1993. Context, Time, and Memory Retrieval in the Interference Paradigms of Pavlovian Learning. Psychol. Bull. 114, 80–99.

Bukalo, O., Pinard, C.R, Silverstein, S., Brehm, C., Hartley, N.D., Whittle, N., Colacicco, G., Busch, E., Patel, S., Singewald, N., Holmes, A. 2015. Prefrontal inputs to the amygdala instruct fear extinction memory formation. Sci.Adv.1 pii: e1500251. https://doi.org/ 10.1126/sciadv.1500251

Ciocchi, S., Herry, C., Grenier, F., Wolff, S.B.E., Letzkus, J.J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M.B., Müller, C., Lüthi, A., 2010. Encoding of conditioned fear in central amygdala inhibitory circuits. Nature 468, 277–282. https://doi.org/10.1038/nature09559

Clem, R.L., Schiller, D., 2016. New Learning and Unlearning: Strangers or Accomplices in Threat Memory Attenuation? Trends Neurosci. 39, 340–351. https://doi.org /10.1016/ j.tins.2016.03.003

Craske, M.G., Treanor, M., Conway, C.C., Zbozinek, T., Vervliet, B., 2014. Maximizing exposure therapy: An inhibitory learning approach. Behav. Res. Ther. 58, 10–23. https://doi.org/10.1016/j.brat.2014.04.006

Dejean, C., Courtin, J., Rozeske, R.R., Bonnet, M.C., Dousset, V., Michelet, T., Herry, C., 2015. Neuronal Circuits for Fear Expression and Recovery: Recent Advances and Potential Therapeutic Strategies. Biol. Psychiatry 78, 298–306. https://doi.org/ 10.1016/ j.biopsych.2015.03.017.

Delgado, M.R., Beer, J.S., Fellows, L.K., Huettel, S.A., Platt, M.L., Quirk, G.J., Schiller, D. 2016. Viewpoints: Dialogues on the functional role of the ventromedial prefrontal Cortex. Nat Neurosci. 19:1545-1552. https://doi.org/10.1038/nn.4438.

Delgado, M.R., Nearing, K.I., LeDoux, J.E., Phelps, E.A., 2008. Neural Circuitry Underlying the Regulation of Conditioned Fear and Its Relation to Extinction. Neuron 59, 829–838. https://doi.org/10.1016/j.neuron.2008.06.029

Diener, S.J., Nees, F., Wessa, M., Wirtz, G., Frommberger, U., Penga, T., Ruttorf, M., Ruf, M., Schmahl, C., Flor, H., 2016. Reduced amygdala responsivity during conditioning to trauma-related stimuli in posttraumatic stress disorder. Psychophysiology 53, 1460–1471. https://doi.org/10.1111/psyp.12699

Diekhof, E.K., Geier, K., Falkai, P., Gruber, O., 2011. Fear is only as deep as the mind allows. A coordinate-based meta-analysis of neuroimaging studies on the regulation of negative affect. Neuroimage 58, 275–285. https://doi.org/10.1016/j.neuroimage.2011.05.073

Do-Monte, F.H., Manzano-Nieves, G., Quinones-Laracuente, K., Ramos-Medina, L., Quirk, G.J., Quiñones-Laracuente, K., Ramos-Medina, L., Quirk, G.J., 2015. Revisiting the role of infralimbic cortex in fear extinction with optogenetics. J. Neurosci. 35, 3607–15. https://doi.org/10.1523/JNEUROSCI.3137-14.2015

Duits, P., Cath, D.C., Lissek, S., Hox, J.J., Hamm, A.O., Engelhard, I.M., Van Den Hout, M.A., Baas, J.M.P., 2015. Updated meta-analysis of classical fear conditioning in the anxiety disorders. Depress. Anxiety 32, 239–253. https://doi.org/10.1002/da.22353

Dunsmoor, J.E., Niv, Y., Daw, N., Phelps, E.A., 2015. Rethinking Extinction. Neuron 88, 47–63. https://doi.org/10.1016/j.neuron.2015.09.028

Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. Br. Med. J. 315, 629–634. https://doi.org/10.1136/bmj.316.7129.469

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. Neuroimage 59, 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp. 30, 2907–2926. https://doi.org/10.1002/hbm.20718

Etkin, A., Egner, T., Kalisch, R., 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. 15, 85–93. https://doi.org/10.1016/j.tics.2010.11.004

Ewald, H., Glotzbach-Schoon, E., Gerdes, A.B.M., Andreatta, M., Müller, M., Mühlberger, A., Pauli, P., 2014. Delay and trace fear conditioning in a complex virtual learning environmentâ€"neural substrates of extinction. Front. Hum. Neurosci. 8, 1–11. https://doi.org/10.3389/fnhum.2014.00323

Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatry 21, 500–508. https://doi.org/10.1038/mp.2015.88

Gershman, S.J., Hartley, C.A., 2015. Individual differences in learning predict the return of fear. Learn. Behav. 43, 243–250. https://doi.org/10.3758/s13420-015-0176-z

Giustino, T.F., Maren, S., 2015. The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. Front. Behav. Neurosci. 9, 298. https://doi.org/10.3389/fnbeh.2015.00298

Gottfried, J.A., Dolan, R.J., 2004. Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. Nat. Neurosci. 7, 1144–1152. https://doi.org/10.1038/nn1314

Harrison, B.J., Fullana, M.A., Soriano-Mas, C., Via, E., Pujol, J., Martínez-Zalacaín, I., Tinoco-Gonzalez, D., Davey, C.G., López-Solà, M., Pérez Sola, V., Menchón, J.M., Cardoner, N., 2015. A neural mediator of human anxiety sensitivity. Hum. Brain Mapp. 36, 3950–3958. https://doi.org/10.1002/hbm.22889

Harrison, B.J., Fullana, M.A., Via, E., Soriano-Mas, C., Vervliet, B., Martínez-Zalacaín, I., Pujol, J., Davey, C.G., Kircher, T., Straube, B., Cardoner, N., 2017. Human ventromedial prefrontal cortex and the positive affective processing of safety signals. Neuroimage 152, 12–18. https://doi.org/10.1016/j.neuroimage.2017.02.080

Harrison, B.J., Pujol, J., Contreras-Rodríguez, O., Soriano-Mas, C., López-Solà, M., Deus, J., Ortiz, H., Blanco-Hinojo, L., Alonso, P., Hernández-Ribas, R., Cardoner, N., Menchón, J.M., 2011. Task-Induced deactivation from rest extends beyond the default mode brain network. PLoS One 6, e22964. https://doi.org/10.1371/journal.pone.0022964

Harrison, B.J., Pujol, J., López-Solà, M., Hernández-Ribas, R., Deus, J., Ortiz, H., Soriano-Mas, C., Yücel, M., Pantelis, C., Cardoner, N., 2008. Consistency and functional specialization in the default mode brain network. Proc. Natl. Acad. Sci. U. S. A. 105, 9781–6. https://doi.org/10.1073/pnas.0711791105

Hartley, C.A., Phelps, E.A., 2010. Changing Fear: The Neurocircuitry of Emotion Regulation. Neuropsychopharmacology 35, 136–146. https://doi.org/10.1038/npp.2009.121

Hauner, K.K., Howard, J.D., Zelano, C., Gottfried, J.A., 2013. Stimulus-specific enhancement of fear extinction during slow-wave sleep. Nat. Neurosci. 16, 1553–1555. https://doi.org/10.1038/nn.3527

Hermann, A., Küpper, Y., Schmitz, A., Walter, B., Vaitl, D., Hennig, J., Stark, R., Tabbert, K.,

2012. Functional Gene Polymorphisms in the Serotonin System and Traumatic Life Events Modulate the Neural Basis of Fear Acquisition and Extinction. PLoS One 7, e44352. https://doi.org/10.1371/journal.pone.0044352

Hermans, D., Craske, M.G., Mineka, S., Lovibond, P.F., 2006. Extinction in Human Fear Conditioning. Biol. Psychiatry. https://doi.org/10.1016/j.biopsych.2005.10.006

Herry, C., Ferraguti, F., Singewald, N., Letzkus, J.J., Ehrlich, I., Lüthi, A., 2010. Neuronal circuits of fear extinction. Eur. J. Neurosci. 31, 599-612. https://doi:10.1111/j.1460-9568.2010.07101.x.

Holt, D.J., Coombs, G., Zeidan, M.A., Goff, D.C., Milad, M.R., 2012. Failure of neural responses to safety cues in schizophrenia. Arch. Gen. Psychiatry 69, 893–903. https://doi.org/10.1001/archgenpsychiatry.2011.2310

Hugdahl, K., 1980. The Three-Systems-Model of Fear and Emotion -a Critical Examination. Behav. Res. Ther. 19, 75–85. https://doi.org/10.1016/0005-7967(81)90114-5

Icenhour, A., Kattoor, J., Benson, S., Boekstegers, A., Schlamann, M., Merz, C.J., Forsting, M., Elsenbruch, S., 2015. Neural circuitry underlying effects of context on human pain-related fear extinction in a renewal paradigm. Hum. Brain Mapp. 36, 3179–3193. https://doi.org/10.1002/hbm.22837

Kalisch, R., 2006. Context-Dependent Human Extinction Memory Is Mediated by a Ventromedial Prefrontal and Hippocampal Network. J. Neurosci. 26, 9503–9511. https://doi.org/10.1523/JNEUROSCI.2021-06.2006

Kattoor, J., Thürling, M., Gizewski, E.R., Forsting, M., Timmann, D., Elsenbruch, S.

2014. Cerebellar contributions to different phases of visceral aversive extinction learning. Cerebellum 13, 1-8. https://doi: 10.1007/s12311-013-0512-9

Krause-Utz, A., Keibel-Mauchnik, J., Ebner-Priemer, U., Bohus, M., Schmahl, C., 2016. Classical conditioning in borderline personality disorder: an fMRI study. Eur. Arch. Psychiatry Clin. Neurosci. 266, 291–305. https://doi.org/10.1007/s00406-015-0593-1

Lange, I., Kasanova, Z., Goossens, L., Leibold, N., De Zeeuw, C.I., van Amelsvoort, T., Schruers, K.. 2015. The anatomy of fear learning in the cerebellum: A systematic meta-analysis. Neurosci Biobehav Rev. 59:83-91. https://doi: 10.1016/ j.neubiorev.2015.09.019

LaBar, K.S., Gatenby, J.C., Gore, J.C., LeDoux, J.E., Phelps, E.A., 1998. Human Amygdala Activation during Conditioned Fear Acquisition and Extinction: a Mixed-Trial fMRI Study. Neuron 20, 937–945. https://doi.org/10.1016/S0896-6273(00)80475-4

LeDoux, J.E., 2014. Coming to terms with fear. Proc. Natl. Acad. Sci. 111, 2871–2878. https://doi.org/10.1073/pnas.1400335111

LeDoux, J.E., Pine, D.S., 2016. Using Neuroscience to Help Understand Fear and Anxiety: A Two-System Framework. Am. J. Psychiatry 173, 1083–1093. https://doi.org/10.1176/appi.ajp.2016.16030353

Lindner, K., Neubert, J., Pfannmöller, J., Lotze, M., Hamm, A.O., Wendt, J., 2015. Fear-potentiated startle processing in humans: Parallel fMRI and orbicularis EMG assessment during cue conditioning and extinction. Int. J. Psychophysiol. 98, 535–545. https://doi.org/10.1016/j.ijpsycho.2015.02.025

Linnman, C., Zeidan, M.A., Furtak, S.C., Pitman, R.K., Quirk, G.J., Milad, M.R., 2012. Resting amygdala and medial prefrontal metabolism predicts functional activation of the fear extinction circuit. Am. J. Psychiatry 169, 415–423. https://doi.org/10.1176/appi.ajp.2011.10121780

Lissek, S., Bradford, D.E., Alvarez, R.P., Burton, P., Espensen-Sturges, T., Reynolds, R.C., Grillon, C., 2014. Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. Soc. Cogn. Affect. Neurosci. 9, 1134–1142. https://doi.org/10.1093/scan/nst096

Lonsdorf, T.B., Haaker, J., Fadai, T., Kalisch, R., 2014. No evidence for enhanced extinction memory consolidation through noradrenergic reuptake inhibition-delayed memory test and reinstatement in human fMRI. Psychopharmacology (Berl). 231, 1949–62. https://doi.org/10.1007/s00213-013-3338-8

Lonsdorf, T.B., Haaker, J., Kalisch, R., 2013. Long-term expression of human contextual fear and extinction memories involves amygdala, hippocampus and ventromedial prefrontal cortex: A reinstatement study in two independent samples. Soc. Cogn. Affect. Neurosci. 9, 1973–1983. https://doi.org/10.1093/scan/nsu018

Lonsdorf, T.B., Menz, M.M., Andreatta, M., Fullana, M.A., Golkar, A., Haaker, J., Heitland, I., Hermann, A., Kuhn, M., Kruse, O., Drexler, S.M., Meulders, A., Nees, F., Pittig, A., Richter, J., Römer, S., Shiban, Y., Schmitz, A., Straube, B., Vervliet, B., Wendt, J., Baas, J.M., Merz, C.J., 2017. Don't fear "fear conditioning": Methodological considerations for the design and analysis of studies on human fear acquisition, extinction, and return of fear. Neurosci. Biobehav. Rev. 77, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026

Lovibond, P.F., 2004. Cognitive Processes in Extinction. Learn. Mem. 11, 495–500. https://doi.org/10.1101/lm.79604

Lovibond, P.F., Shanks, D.R., 2002. The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. J. Exp. Psychol. Anim. Behav. Process. 28, 3–26. https://doi.org/10.1037/0097-7403.28.1.3

Lueken, U., Straube, B., Reinhardt, I., Maslowski, N.I., Wittchen, H.-U., Ströhle, A., Wittmann, A., Pfleiderer, B., Konrad, C., Ewert, A., Uhlmann, C., Arolt, V., Jansen, A., Kircher, T., 2014. Altered top-down and bottom-up processing of fear conditioning in panic disorder with agoraphobia. Psychol. Med. 44, 381–394. https://doi.org/10.1017/S0033291713000792

Maren, S., 2014. Nature and causes of the immediate extinction deficit: A brief review. Neurobiol. Learn. Mem. https://doi.org/10.1016/j.nlm.2013.10.012

Maren, S., Phan, K.L., Liberzon, I., 2013. The contextual brain: implications for fear conditioning, extinction and psychopathology. Nat. Rev. Neurosci. 14, 417–28. https://doi.org/10.1038/nrn3492

Marin, M.-F., Song, H., Vanelzakker, M.B., Staples-Bradley, L.K., Linnman, C., Pace-Schott, E.F., Lasko, N.B., Shin, L.M., Milad, M.R., 2016. Association of Resting Metabolism in the Fear Neural Network With Extinction Recall Activations and Clinical Measures in Trauma-Exposed Individuals. Am. J. Psychiatry 1–9. https://doi.org/10.1176/appi.ajp.2015.14111460

Mechias, M.L., Etkin, A., Kalisch, R., 2010. A meta-analysis of instructed fear studies: Implications for conscious appraisal of threat. Neuroimage 49, 1760–1768. https://doi.org/10.1016/j.neuroimage.2009.09.040

Medford, N., Critchley, H.D., 2010. Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. Brain Struct. Funct. 214, 535–549. https://doi.org/10.1007/s00429-010-0265-x

Menz, M.M., Rihm, J.S., Buchel, C., 2016. REM Sleep Is Causal to Successful Consolidation of Dangerous and Safety Stimuli and Reduces Return of Fear after Extinction. J. Neurosci. 36, 2148–2160. https://doi.org/10.1523/JNEUROSCI.3083-15.2016

Menz, M.M., Rihm, J.S., Salari, N., Born, J., Kalisch, R., Pape, H.C., Marshall, L., Büchel, C., 2013. The role of sleep and sleep deprivation in consolidating fear memories. Neuroimage 75, 87–96. https://doi.org/10.1016/j.neuroimage.2013.03.001

Merz, C.J., Hermann, A., Stark, R., Wolf, O.T., 2014. Cortisol modifies extinction learning of recently acquired fear in men. Soc. Cogn. Affect. Neurosci. 9, 1426–1434. https://doi.org/10.1093/scan/nst137

Merz, C.J., Tabbert, K., Schweckendiek, J., Klucken, T., Vaitl, D., Stark, R., Wolf, O.T., 2012. Neuronal correlates of extinction learning are modulated by sex hormones. Soc. Cogn. Affect. Neurosci. 7, 819–830. https://doi.org/10.1093/scan/nsr063

Milad, M.R., Furtak, S.C., Greenberg, J.L., Keshaviah, A., Im, J.J., Falkenstein, M.J., Jenike, M., Rauch, S.L., Wilhelm, S., 2013. Deficits in Conditioned Fear Extinction in Obsessive-Compulsive Disorder and Neurobiological Changes in the Fear Circuit. JAMA psychiatry 70, 608. https://doi.org/10.1001/jamapsychiatry.2013.914

Milad, M.R., Quirk, G.J., 2012. Fear extinction as a model for translational neuroscience: ten years of progress. Annu. Rev. Psychol. 63, 129–51. https://doi.org/10.1146/annurev.psych.121208.131631

Milad, M.R., Wright, C.I., Orr, S.P., Pitman, R.K., Quirk, G.J., Rauch, S.L., 2007. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biol. Psychiatry 62, 446–54. https://doi.org/10.1016/j.biopsych.2006.10.011

Molapour, T., Golkar, A., Navarrete, C.D., Haaker, J., Olsson, A., 2015. Neural correlates of biased social fear learning and interaction in an intergroup context. Neuroimage 121, 171–183. https://doi.org/10.1016/j.neuroimage.2015.07.015

Morris, J.S., Buchel, C., Dolan, R.J., 2001. Parallel Neural Responses in Amygdala Subregions and Sensory Cortex during Implicit Fear Conditioning. Neuroimage 13, 1044–1052. https://doi.org/10.1006/nimg.2000.0721

Morriss, J., Christakou, A., van Reekum, C.M., 2015. Intolerance of uncertainty predicts fear extinction in amygdala-ventromedial prefrontal cortical circuitry. Biol. Mood Anxiety Disord. 5, 4. https://doi.org/10.1186/s13587-015-0019-8

Myers, K.M., Davis, M., 2002. Behavioral and neural analysis of extinction. Neuron 36, 567–584. https://doi.org/10.1016/S0896-6273(02)01064-4

Myers, K.M., Ressler, K.J., Davis, M., 2006. Different mechanisms of fear extinction dependent on length of time since fear acquisition. Learn. Mem. 13, 216–223. https://doi.org/10.1101/lm.119806

Norrholm, S.D., Vervliet, B., Jovanovic, T., Boshoven, W., Myers, K.M., Davis, M., Rothbaum, B., Duncan, E.J., 2008. Timing of extinction relative to acquisition: a parametric analysis of fear extinction in humans. Behav. Neurosci. 122, 1016–1030. https://doi.org/10.1037/a0012604

Ochsner, K.N., Gross, J.J., 2005. The cognitive control of emotion. Trends Cogn. Sci. 9, 242–249. https://doi.org/10.1016/j.tics.2005.03.010

Pavlov, I.P., 1929. Conditioned Reflexes. An investigation of the physiological activity of the cerebral cortex. Oxford University Press.

Pejic, T., Hermann, A., Vaitl, D., Stark, R., 2013. Social anxiety modulates amygdala activation during social conditioning. Soc. Cogn. Affect. Neurosci. 8, 267–276. https://doi.org/10.1093/scan/nsr095

Pennington, Z.T., Anderson, A.S., Fanselow, M.S. 2017. The ventromedial prefrontal cortex in a model of traumatic stress: fear inhibition or contextual processing? Learn. Mem. 16, 400-406. https://doi.org/10.1101/lm.046110.117

Phelps, E.A., Delgado, M.R., Nearing, K.I., Ledoux, J.E., 2004. Extinction learning in humans: Role of the amygdala and vmPFC. Neuron 43, 897–905. https://doi.org/10.1016/j.neuron.2004.08.042

Quirk, G.J., Mueller, D., 2008. Neural mechanisms of extinction learning and retrieval. Neuropsychopharmacology 33, 56–72. https://doi.org/10.1038/sj.npp.1301555

Rabinak, C.A., Angstadt, M., Lyons, M., Mori, S., Milad, M.R., Liberzon, I., Phan, K.L., Luan Phan, K., 2014. Cannabinoid modulation of prefrontalâ€“limbic activation during fear extinction learning and recall in humans. Neurobiol. Learn. Mem. 113, 125–134. https://doi.org/10.1016/j.nlm.2013.09.009

Repa, J.C., Muller, J., Apergis, J., Desrochers, T.M., Zhou, Y., LeDoux, J.E. 2001. Two different lateral amygdala cell populations contribute to the initiation and storage of memory. Nat.

Neurosci. 4:724-31. https://doi.org/10.1038/89512

Raczka, K., Mechias, M.-L., Gartmann, N., Reif, A., Deckert, J., Pessiglione, M., Kalisch, R., 2011. Empirical support for an involvement of the mesostriatal dopamine system in human fear extinction. Transl. Psychiatry 1. https://doi.org/10.1038/tp.2011.10

Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., Surguladze, S., 2012. A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. Eur. Psychiatry 27, 605–611. https://doi.org/10.1016/j.eurpsy.2011.04.001

Radua, J., Rubia, K., Canales-Rodríguez, E.J., Pomarol-Clotet, E., Fusar-Poli, P., Mataix-Cols, D., 2014. Anisotropic kernels for coordinate-based meta-analyses of neuroimaging studies. Front. Psychiatry 5. https://doi.org/10.3389/fpsyt.2014.00013

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 676–82. https://doi.org/10.1073/pnas.98.2.676

Reijmers, L.G., Perkins, B.L., Matsuo, N., Mayford, M., 2007. Localization of a stable neural correlate of associative memory. Science 317, 1230–3. https://doi.org/10.1126/science.1143839

Reinhardt, I., Jansen, A., Kellermann, T., Schüppen, A., Kohn, N., Gerlach, A.L., Kircher, T., 2010. Neural correlates of aversive conditioning: development of a functional imaging paradigm for the investigation of anxiety disorders. Eur. Arch. Psychiatry Clin. Neurosci. 260, 443–453. https://doi.org/10.1007/s00406-010-0099-9

Rescorla, R., Wagner, A.., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement, in: Black, A.H., Prokasky, W.. (Eds.), Classical Conditioniing !!: Current Research and Theory. Appleton-Century-Crofts, New York, pp. 64–99.

Ridder, S., Treutlein, J., Nees, F., Lang, S., Diener, S., Wessa, M., Kroll, A., Pohlack, S., Cacciaglia, R., Gass, P., Schütz, G., Schumann, G., Flor, H., 2012. Brain activation during fear conditioning in humans depends on genetic variations related to functioning of the hypothalamic–pituitary–adrenal axis: first evidence from two independent subsamples. Psychol. Med. 42, 2325–2335. https://doi.org/10.1017/S0033291712000359

Rothbaum, B.O., Kearns, M.C., Price, M., Malcoun, E., Davis, M., Ressler, K.J., Lang, D., Houry, D., 2012. Early intervention may prevent the development of posttraumatic stress disorder: A randomized pilot civilian study with modified prolonged exposure. Biol. Psychiatry 72, 957–963. https://doi.org/10.1016/j.biopsych.2012.06.002

Schiller, D., Delgado, M.R., 2010. Overlapping neural systems mediating extinction, reversal and regulation of fear. Trends Cogn. Sci. 14, 268–276. https://doi.org/10.1016/j.tics.2010.04.002

Schiller, D., Levy, I., Niv, Y., LeDoux, J.E., Phelps, E.A., 2008. From Fear to Safety and Back: Reversal of Fear in the Human Brain. J. Neurosci. 28, 11517–11525. https://doi.org/10.1523/JNEUROSCI.2265-08.2008

Sehlmeyer, C., Dannlowski, U., Schöning, S., Kugel, H., Pyka, M., Pfleiderer, B., Zwitserlood, P., Schiffbauer, H., Heindel, W., Arolt, V., Konrad, C., 2011. Neural correlates of trait anxiety in fear extinction. Psychol. Med. 41, 789–798.

https://doi.org/10.1017/S0033291710001248

Shumake, J., Furgeson-Moreira, S., Monfils, M.H., 2014. Predictability and heritability of individual differences in fear learning. Anim. Cogn. 17, 1207–1221. https://doi.org/10.1007/s10071-014-0752-1

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283, 2008–12.

Tovote, P., Fadok, J.P., Lüthi, A., 2015. Neuronal circuits for fear and anxiety. Nat. Rev. Neurosci. 16, 317–331. https://doi.org/10.1038/nrn3945

Utz, A., Thürling, M., Ernst, T.M., Hermann, A., Stark, R., Wolf, O.T., Timmann, D., Merz, C.J. 2015. Cerebellar vermis contributes to the extinction of conditioned fear. Neurosci Lett. 604:173-7. https://doi: 10.1016/j.neulet.2015.07.026

Vervliet, B., Craske, M.G., Hermans, D., 2013. Fear Extinction and Relapse: State of the Art. Annu. Rev. Clin. Psychol. 9, 215–248. https://doi.org/10.1146/annurev-clinpsy-050212-185542

Wicking, M., Steiger, F., Nees, F., Diener, S.J., Grimm, O., Ruttorf, M., Schad, L.R., Winkelmann, T., Wirtz, G., Flor, H., 2016. Deficient fear extinction memory in posttraumatic stress disorder. Neurobiol. Learn. Mem. 136, 116–126. https://doi.org/10.1016/j.nlm.2016.09.016

FIGURE LEGENDS

Figure 1: Neural correlates of fear extinction learning estimated by meta-analysis. Results are displayed at $p < 0.005$ (cluster size ≥10 voxels) on the MNI 152 T1 0.5mm template.

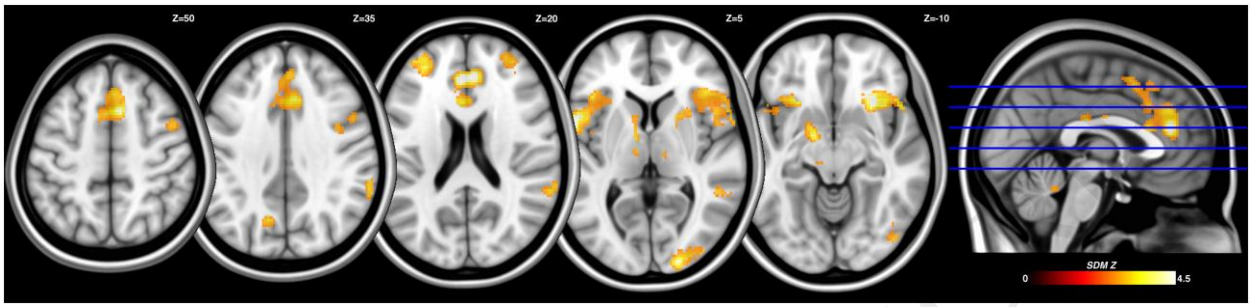

Figure 2: Neural correlates of fear conditioning versus extinction learning estimated by meta-analysis (conditioning > extinction). Results are displayed at $p < 0.005$ (cluster size ≥10 voxels) on the MNI 152 T1 0.5mm template.

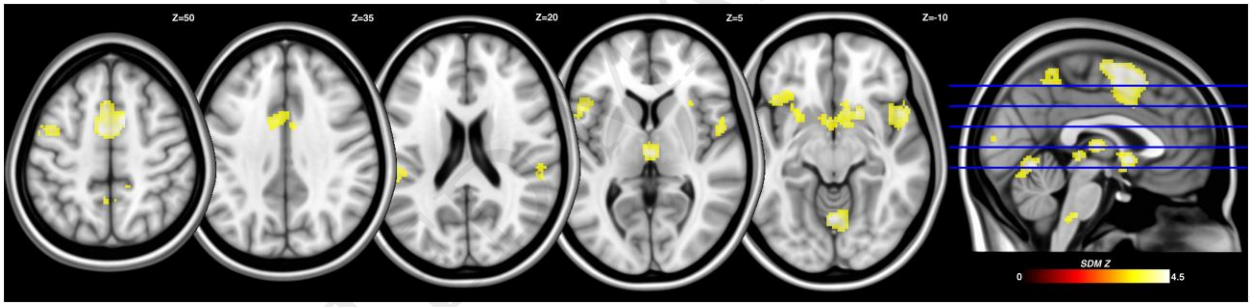

Figure 3: Neural correlates of fear extinction recall estimated by meta-analysis. A: Extinction recall overall results, B: Extinction recall CS+ > CS- contrast results. Results are displayed at p < 0.005 (cluster size ≥10 voxels) on the MNI 152 T1 0.5mm template.

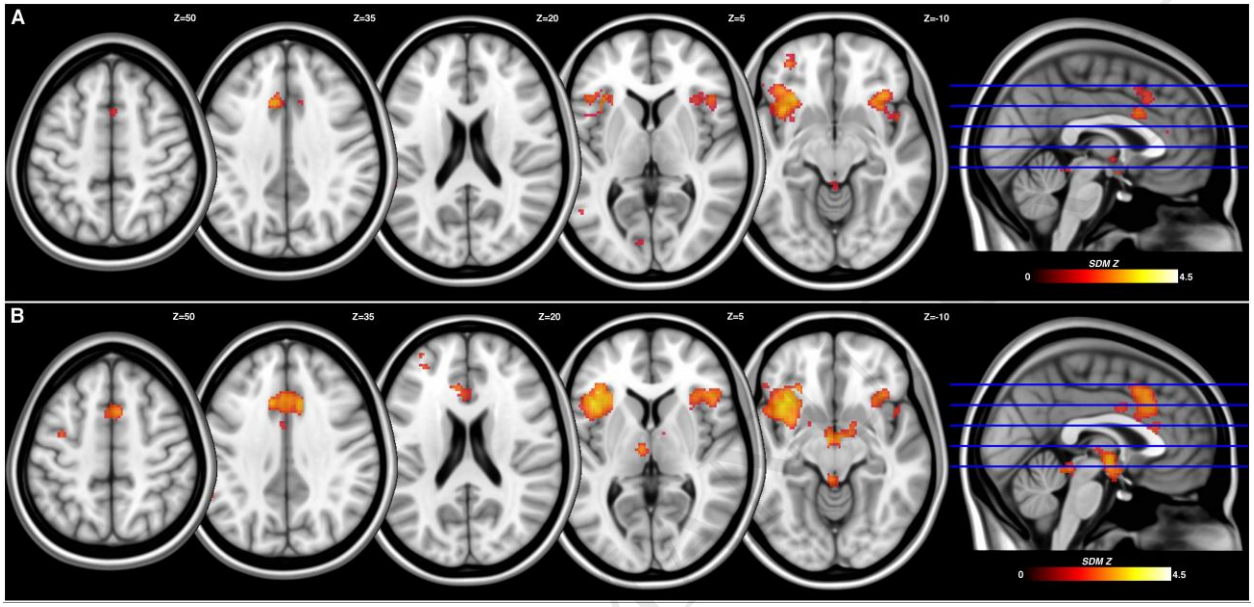

Figure 4: Neural correlates of fear extinction recall estimated by meta-analysis (CS+E > CS+U contrast results). Results are displayed at p < 0.005 (cluster size ≥10 voxels) on the MNI 152 T1 0.5mm template.

Fear extinction in the human brain

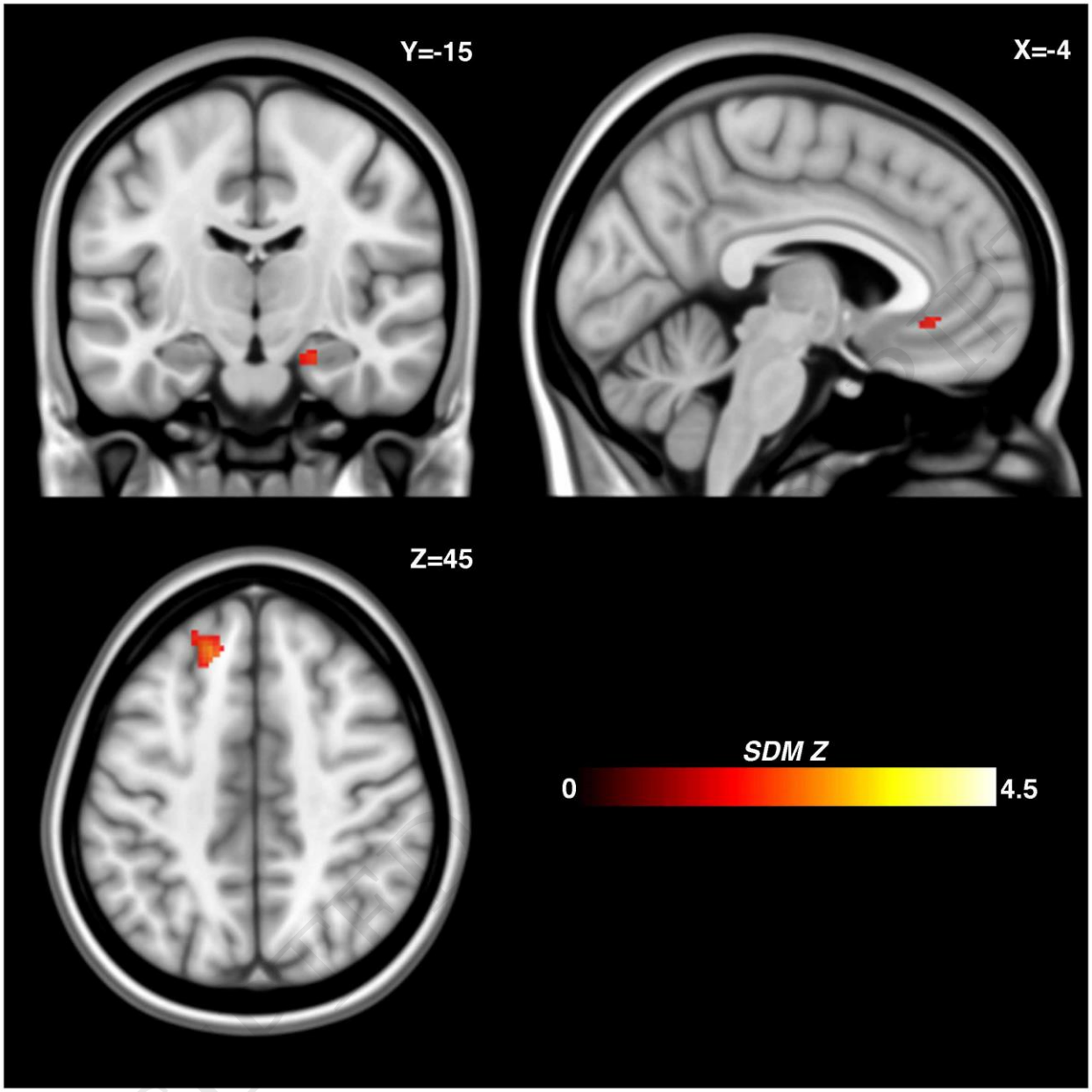

Table 1. Characteristics of the 31 fMRI studies included in the extinction learning meta-analysis.

Authors	N	Males (%)	Mean age (years)	CS	Reinforcement rate during conditioning (%)	Change in context from conditioning to extinction	Immediate extinction?	Number of CS+/CS- during conditioning	Number of CS+/CS- during extinction	fMRI analysis [a]
Ahs et al. 2015	43	49	28.7	Dynamic images (VR)	31	YES	YES	16/16	16/16	Early, late *
Benson et al. 2014	29	50	23.8	Geometrical figures	75	NO	YES	16/16	12/12	Early, late *
Diener et al. 2016	13	53	42.46	Geometrical figures	50	NO	YES	18/18	18/18	Whole *
Ewald et al. 2014	13	38	23.1	Lights (VR)	100	NO	YES	16/16	8/8	Early, late
Harrison et al., unpublished	58	33	21.8	Geometrical figures	50	NO	YES	32/ 32	16/16	Whole, early, late *
Hermann et al.2012	74	50	24.3	Geometrical figures	100	NO	YES	20/20	15/15	Whole *
Holt et al.2012	17	100	34.2	Photographs	60	YES	YES	16/16	16/16	Early
Icenhour et al.2015	23	48	33.7	Geometrical figures	75	NO	YES	16/16	6/6	Early, late *
Klumpers et al.unpublished	106	100	21.9	Geometrical figures	33	NO	YES	18/18	18/18	Whole, early, late *
Krause-Utz et al. 2015	26	0	28.16	Geometrical figures	50	NO	YES	36/36	18/18	Whole
Kuhn et al., unpublished	37	49	25.13	Geometrical figures	100	NO	YES	18/18	12/12	Whole *
Lindner et al.2015	15	0	22.53	Geometrical figures	100	NO	YES	8/8	4/4	Whole *
Linmann et al.2012	18	44	25.7	Photographs	62	YES	YES	16/16	16/16	Early, late
Lonsdorf et al. 2014	59	46	24	Angry faces	100	NO	NO	15/15	24/24	Whole *
Lueken et al. 2014	60	32	35.75	Geometrical figures	50	NO	YES	32/32	16/16	Whole, early, late *
Merz et al. 2012	49	41	24.33	Geometrical figures	100	NO	YES	20/20	11/11	Whole *
Merz et al. 2014	16	100	24.88	Geometrical figures	62	NO	YES	16/16	16/16	Early, late *
Milad et al. 2007	14	NA	NA	Photographs	60	YES	YES	16/16	16/16	Late
Milad et al. 2013	16	NA	25.8	Photographs	62	YES	YES	16/16	16/16	Early vs late
Molapour et al. 2015	20	50	22.39	Neutral faces	100	NO	YES	9/9	12/12	Whole *
Morris et al. 2015	21	48	24.03	Geometrical figures	100	NO	YES	12/12	16/16	Whole *
Pejic et al. 2013	49	54	23.49	Neutral faces	100	NO	YES	17/17	2x(13/13)	Whole *
Phelps et al. 2004	11	45	NA	Geometrical figures	33	NO	YES	23/15	15/15	Whole *
Rabinak et al. 2014	14	64	25.43	Geometrical figures	35	YES	YES	23-23/15	30/30	Early, late

Reinhardt et al. 2010	20	100	28.8	Geometrical figures	50	NO	YES	32/16	16/16	Whole
Ridder et al. 2012, sample 1	60	63	21.25	Geometrical figures	50	NO	YES	18/18	18/18	Whole
Scharfenort et al.unpublished	77	47	24.8	Geometrical figures	100	NO	NO	14/14	14/14	Whole *
Sehlmeyer et al. 2011	32	38	23.6	Neutral faces	25	NO	YES	40/30	25/25	Whole
Soriano-Mas et al. unpublished	18	55	35.6	Photographs	62	YES	YES	16/16	16/16	Whole
Spoormaker et al. unpublished	48	87	24.9	Geometrical figures	50	NO	COMBINED	30/15	15/15	Whole *
Wicking et al. 2016	18	61	38.6	Geometrical figures	100	YES	NO	30/30	30/30	Early, late *
TOTAL/mean	**1074**	**61 %**	**27**		**69%**					

Abbreviations: CS, conditioned stimulus; CS+, CS followed by unconditioned stimulus; CS −, CS not followed by unconditioned stimulus, fMRI=functional magnetic resonance imaging, NA= Not available, VR= Virtual Reality

* Datasets for which statistical parametric maps were available.

[a] Contrast available for *within-session* fear extinction: whole (whole extinction), early (early extinction trials) and late (late extinction trials).

Table 2. Characteristics of the 16 fMRI studies included in the extinction recall meta-analysis.

Authors	N	Males (%)	Mean age (years)	CS	Interval *extinction learning/extinction recall*	Extinction recall context	fMRI contrast
Ahs et al.2015	43	49	28.7	Dynamic images (VR)	24 hours	Same as extinction learning	CS+ vs CS- *
Holt et al.2012	17	100	34.2	Photographs	24 hours	Same as extinction learning	CS+E vs CS+U
Kalisch et al. 2006	17	53	25	Angry faces	24 hours	Same as extinction learning	CS+ vs CS-
Krause-Utz et al.2015	22	0	27.95	Geometrical figures	72 hours	Same as conditioning	CS+ vs CS-
Linmann et al. 2012	18	44	25.7	Photographs	24 hours	Same as extinction learning	CS+E vs CS+U
Lonsdorf 2013 et al. sample 1	20	100	28.8	Geometrical figures	7 days	Different to conditioning/extinction learning [a]	CS+ vs CS- *
Lonsdorf 2013 et al. sample 2	19	100	29.2	Geometrical figures	7 days	Different to conditioning/extinction learning [a]	CS+ vs CS- *
Menz et al. 2016	20	0	26.55	Geometrical figures	30 hours	Same as extinction learning	CS+E vs CS+U *
Menz et al. 2013	20	0	23.95	Geometrical figures	36 hours	Same as extinction learning	CS+E vs CS+U *
Menz et al. unpublished	21	0	25.68	Geometrical figures	30 hours	Same as extinction learning	CS+E vs CS+U *
Milad et al. 2007	14	NA	NA	Photographs	24 hours	Same as extinction learning	CS+E vs CS+U
Milad et al.2013	19	NA	NA	Photographs	24 hours	Same as extinction learning	CS+E vs CS+U
Pejic et al. 2013	49	54	23.49	Neutral faces	24 hours	Same as conditioning	CS+ vs CS- *
Phelps et al. 2004	11	45	NA	Geometrical figures	24 hours	Same as conditioning	CS+ vs CS-
Rabinak et al. 2014	14	64	25.43	Geometrical figures	24 hours	Same as extinction learning	CS+E vs CS+U
Soriano-Mas et al. unpublished	18	55	35.6	Photographs	24 hours	Same as extinction learning	CS+E vs CS+U *
TOTAL/mean	**342**	**47%**	**27.4**				

Abbreviations: CS, conditioned stimulus; CS+, CS followed by unconditioned stimulus; CS −, CS not followed by unconditioned stimulus; CS+E, extinguished CS+, CS+U, unextinguished CS+, fMRI=functional magnetic resonance imaging, NA= Not available, VR= Virtual Reality.
* Datasets for which statistical parametric maps were available.
[a] Conditioning and extinction learning conducted outside the scanner.